Intro to Linux

1.2.2 - File Compressing and Archiving

CYBER.ORG
THE ACADEMIC INITIATIVE OF THE CYBER INNOVATION CENTER

File Compressing and Archiving

- Compressing and archiving are fundamental processes for Linux users
- Optimize disk space utilization
  - Size of data is reduced allowing for efficient storage and faster transmission
- Ensure safety and integrity of data
  - Safeguard against loss or deletion
  - Allow restoration in the event of a hardware failure or malicious attack

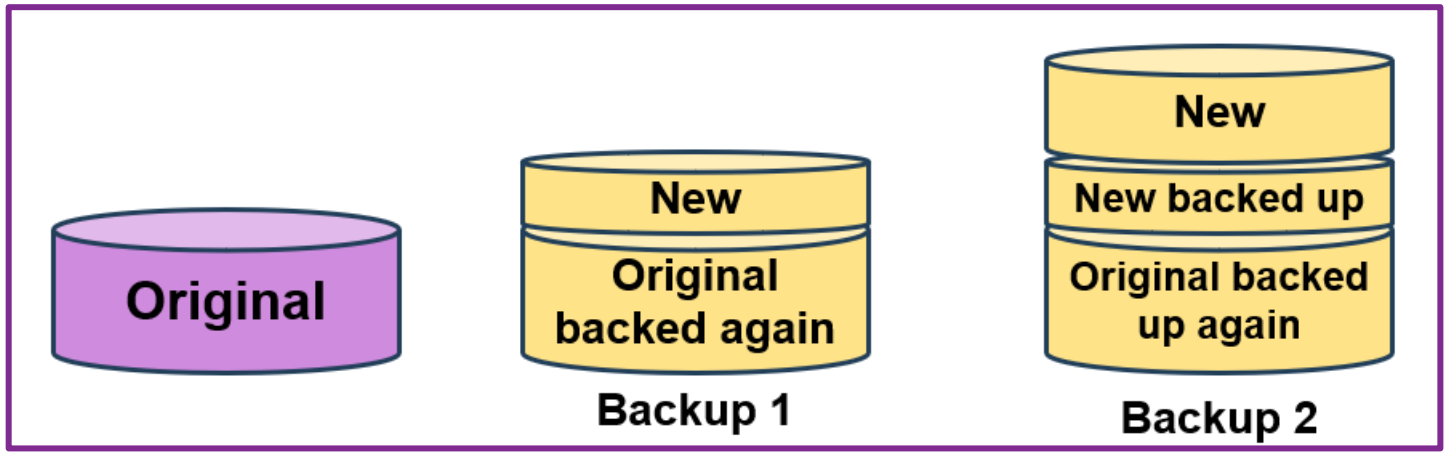
System Images

- Clones of the operating system and configuration
- Do not recover directories or files but allow a fast bootup for a system

Backup and Backup Types

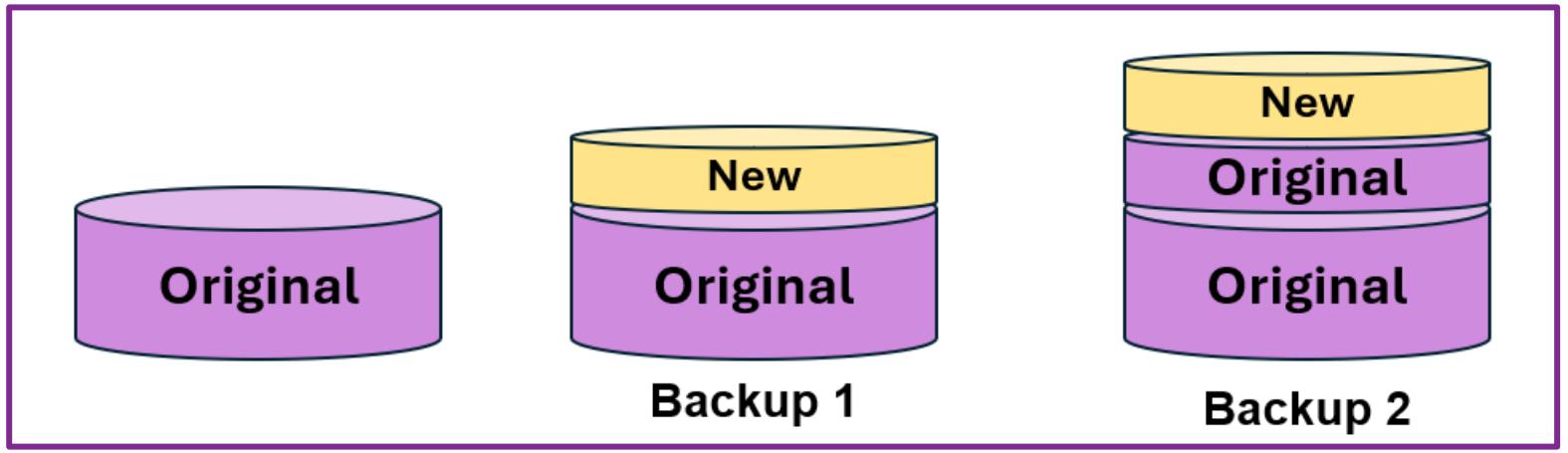
Backups, occasionally referred to as archives, aid in restoring data that has been compromised

- Full Backup
  - Copies everything on the system
  - Useful for recovering corrupted or lost files
  - Very time consuming
  - Takes up a ton of space



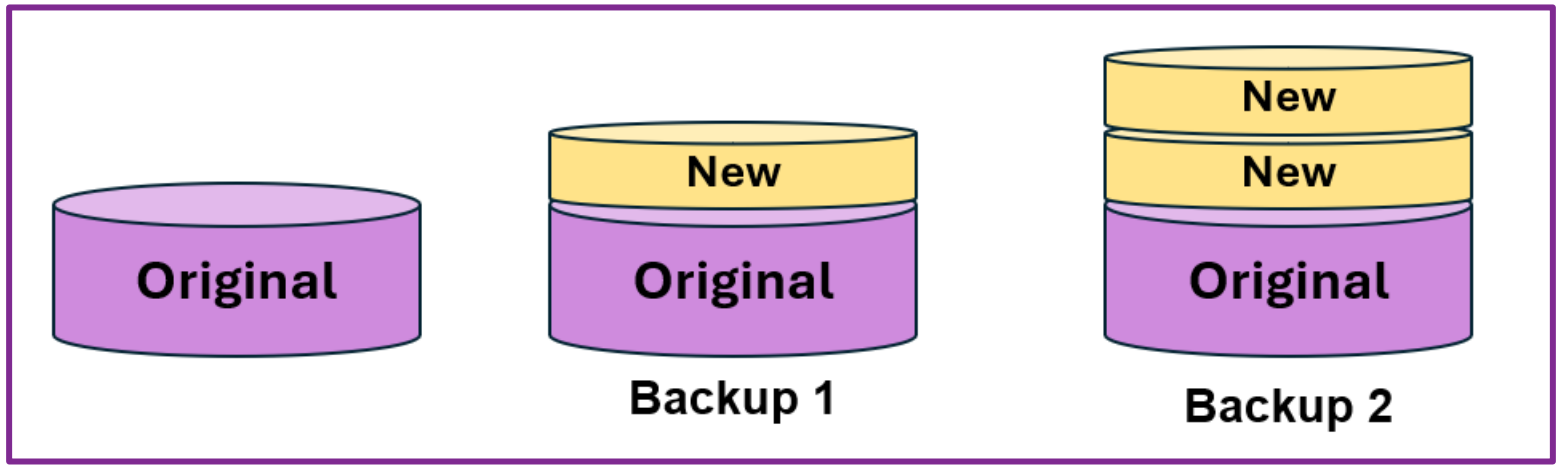
Backup Types - Incremental

- Incremental Backup
  - Copies everything on the system only if it has been modified since the last backup by cross checking timestamps
  - Takes less time that a full backup
  - Each backup takes up less storage
  - Recovery can be very time consuming due to having multiple components to pull from



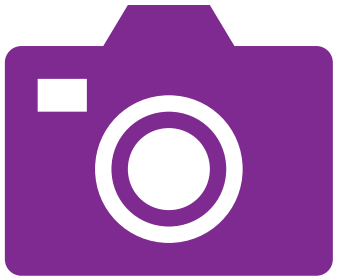
Backup Types - Differential

- Differential Backup
  - Like an incremental backup in that it only backs up what has been modified but differs in that it goes by the timestamp of the most recent full backup
  - Occurs less often that an incremental backup
  - Stronger chance that data could be lost



Snapshots

- Creates a full backup that is read-only
- Takes a “snapshot” of the directory and files’ metadata and how things are stored
- The read-only aspect makes the snapshot much faster than traditional backups
- Snapshots of the metadata can occur multiple times a day



Archiving

- Like backups in that a copy is typically made
- Usually limited in size or specific to a set of data
- Often used with sensitive data, or data that must be kept for long periods of time such as financials but may not need accessing on a regular basis

Archiving – cpio Utility

- Stands for copy in and out
- Can be run within a directory to archive those specific files
- **`cpio`** maintains the directory's references and outputs the data plus the files and directories contained into one archive file

```
ubuntu@ip-10-15-86-128:~/Documents$ pwd
/home/ubuntu/Documents
ubuntu@ip-10-15-86-128:~/Documents$ ls
ubuntu@ip-10-15-86-128:~/Documents$ ls | cpio -o > ArchivedFiles.cpio
1 block
ubuntu@ip-10-15-86-128:~/Documents$ ls
ArchivedFiles.cpio
ubuntu@ip-10-15-86-128:~/Documents$
```

Archiving – tar Archiver

- Stands for tape archiver
- Similar to **`cpio`** but has the option to compress the files as well
- Uncompressed data is referred to as a tar archive file
- Compressed data is referred to as a tarball

```
ubuntu@ip-10-15-86-128:~/Documents$ ls
ArchivedFiles.cpio  file1
ubuntu@ip-10-15-86-128:~/Documents$ tar -zcf Archive.tar.gz file?
ubuntu@ip-10-15-86-128:~/Documents$ ls
Archive.tar.gz  ArchivedFiles.cpio  file1
```

File Compression – gzip, bzip2, and xz

- All of the utilities compress files using the zip and unzip mechanism
- The commands are the name of the utility with the file name as seen with **`gzip file1`**
- bzip compresses files more than gzip but takes longer
- xz can compress at a higher rate but also offers options to change the rate

```
ubuntu@ip-10-15-86-128:~/Documents$ ls
Archive.tar.gz  ArchivedFiles.cpio  file1  file1.gz  file2  file3
ubuntu@ip-10-15-86-128:~/Documents$ rm file1.gz
ubuntu@ip-10-15-86-128:~/Documents$ LS
LS: command not found
ubuntu@ip-10-15-86-128:~/Documents$ ls
Archive.tar.gz  ArchivedFiles.cpio  file1  file2  file3
ubuntu@ip-10-15-86-128:~/Documents$ gzip file1
ubuntu@ip-10-15-86-128:~/Documents$ ls
Archive.tar.gz  ArchivedFiles.cpio  file1.gz  file2  file3
ubuntu@ip-10-15-86-128:~/Documents$ bzip2 file2
ubuntu@ip-10-15-86-128:~/Documents$ ls
Archive.tar.gz  ArchivedFiles.cpio  file1.gz  file2.bz2  file3
ubuntu@ip-10-15-86-128:~/Documents$ xz file3
ubuntu@ip-10-15-86-128:~/Documents$ ls
Archive.tar.gz  ArchivedFiles.cpio  file1.gz  file2.bz2  file3.xz
ubuntu@ip-10-15-86-128:~/Documents$
```

Backing Up Entire Drives with dd

- The dd utility allows a user to copy the contents of a disk or partition onto a new drive or partition
- Requires some additional steps such as ensuring the drive(s) being copied and the one(s) being copied to are not mounted, which may involve booting from an external source